Parameter selection for segregating speech from background noise

THE OHIO STATE UNIVERSITY

Jordan L. Vasko, Sarah E. Yoho, Eric W. Healy

Department of Speech and Hearing Science

Purpose:

- To determine improvements in phoneme recognition for normal hearing (NH) listeners resulting from parameter changes to the ideal binary mask (IBM)
- To determine the optimal relative criterion (RC) for NH listeners hearing IBM-processed speech, in the absence of a ceiling effect.
- Results may be used to determine (1) parameters for adjusting the IBM to account for the contribution of different frequency bands to speech understanding, and (2) parameters in future algorithms to estimate the IBM

Method:

- **IBM Processing:**

  Prior knowledge of speech signal and noise background

  64-channel gammatone filterbank

  20-ms bins, 10-ms overlap

  Calculated SNR in each Time-Frequency (T-F) unit

  Mask Generation:

$$\text{IBM(t,f)} = \begin{cases} 1, \text{ if SNR (t,f)} > \text{Local Criterion (LC)} & (unit\ retained) \\ 0, \text{ otherwise} & (unit\ discarded) \end{cases}$$

  IBM speech = speech+noise mixture in 1-valued T-F units

- **Subjects and Stimuli**

  10 young NH listeners, repeated-measures design

  16 /aCa/ consonants, 4 male talkers

  Background: Speech-shaped noise (SSN) at -8dB SNR

  RC = LC – Input SNR

  Conditions: 7 LC values to yield 7 RC values (see Table 1)

  Presentation level: 65 dBA (comfortable conversational level)

Table 1. Conditions

LC (dB SNR)	Input SNR (dB SNR)	RC (dB SNR) [=LC - Input SNR]	Percent Ones*
-28	-8	**-20**	79.83%
-23	-8	**-15**	69.10%
-18	-8	**-10**	55.65%
-13	-8	**-5**	37.07%
-8	-8	**0**	17.00%
-3	-8	**5**	7.24%
2	-8	**10**	2.56%

*averaged across the Talker 1 masks for all 16 consonants

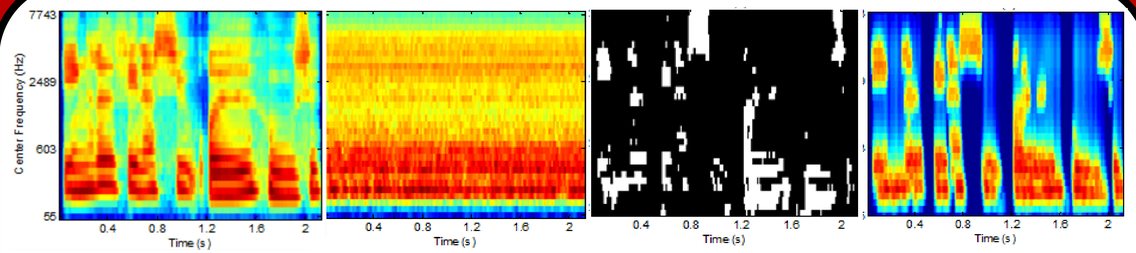

Fig 1. 32-channel cochleagrams of normal speech, speech-shaped noise, ideal binary mask (IBM), and speech-noise mixture after IBM is applied. From D.L. Wang., U. Kjems, M.S. Pedersen, J.B. Boldt, and T. Lunner, J. Acoust. Soc. Am. 124(4), 2303-2307 (2008).

1. Previous RC Performance Functions

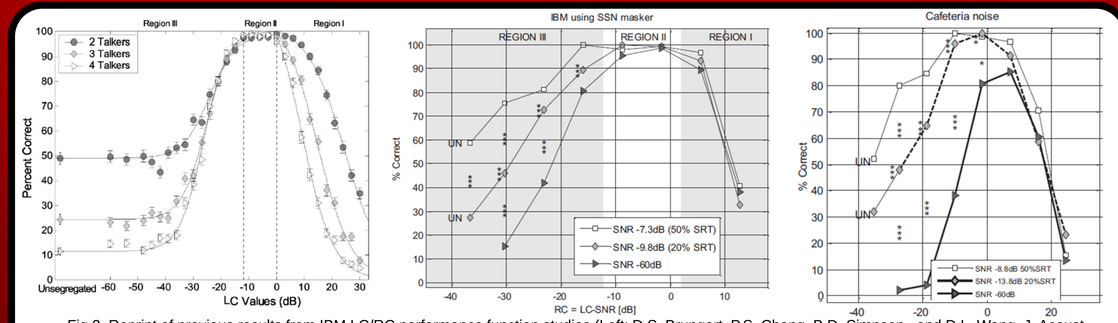

Fig 2. Reprint of previous results from IBM LC/RC performance function studies (Left: D.S. Brungart, P.S. Chang, B.D. Simpson, and D.L. Wang, J. Acoust. Soc. Am. 120(6), 4007-4018 (2006), CRM phrases in multitalker babble at 0 dB SNR; Middle and Right: U. Kjems, J.B. Boldt, M.S. Pedersen, T. Lunner, and D.L. Wang, J. Acoust. Soc. Am. 126(3), 1415-1426 (2009); Dantale II sentences in SSN and cafeteria noise, respectively, at various SNRs).

2a. Current RC Performance Function

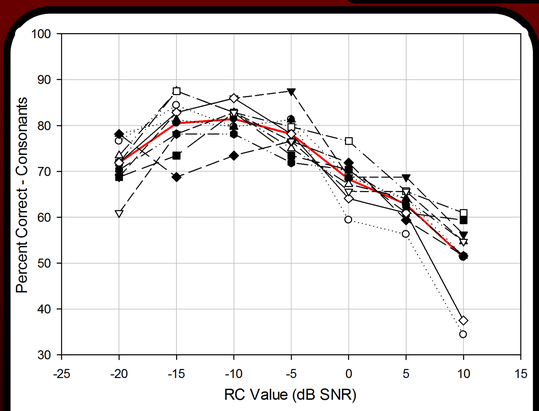

Fig 3. VCV performance for the 10 listeners across conditions. Each black line represents a separate listener, and the red line with filled circles represents the mean across subjects.

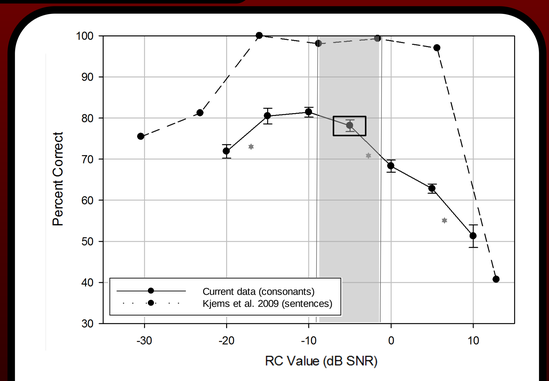

Fig 4. Mean RC performance function (with SEs) for the current study and for Kjems et al. (2009). The gray region is the plateau proposed by Kjems et al. For the current data, asterisks(*) indicate statistically significant differences between adjacent conditions. The box indicates the RC values used in the IBM estimation algorithms (e.g. E.W. Healy, S.E. Yoho, Y. Wang, and D.L. Wang, J. Acoust. Soc. Am. 134(4), 3029-3038 (2013).)

2b. Pairwise Comparisons between RC Conditions

Table 2. Bonferroni Post hoc Results

RC Condition	Signficantly Different from*	Not Significantly Different from*
-20	**-15,** -10, 5, 10	0, -5
-15	**-20**, 0, 5, 10	**-10**, **-5**
-10	-20, 0, 5, 10	**-15**, **-5**
-5	**0**, 5, 10	-20, **-15**, **-10**
0	-15, -10, **-5**, 10	-20, **5**
5	-20, -15, -10, -5, **10**	0
10	-20, -15, -10, -5, 0, **5**	n/a

*Pairwise comparisons are significantly different at $p < 0.0024$ for familywise $\alpha = .050$ with 21 comparisons. One-way RM ANOVA was significant with $F(6,54) = 43.873$, $p < 0.001$. Bolded values represent comparisons between adjacent conditions.

Conclusions and Future Directions:

Conclusions

- The roughly 10-dB wide performance plateau largely remains in absence of a ceiling effect, but occurs at more negative RC values than previously indicated. This reflects an increased tolerance for noise in the masked output, in favor of retaining more of the original mixture.
- The plateau shift could result from: (1) the current use of a more representative SNR; (2) the current use of consonants rather than sentences, as previously used. There could potentially be an effect of speech material.

Future directions

- *(A):* Compare performance with a “variable-LC” mask against the fixed-LC mask (with RC ϵ[-15, -10, -5] dB SNR and an overall input SNR of -8 dB). The current data provide an indication of the optimal fixed-LC value, and guidance for the selection of variable LCs.
- *(B):* Repeat the study to find an RC performance function for hearing impaired participants. Different results may be expected given their decreased tolerance for noise.
- *(C):* Test the effect of a lower RC value for masks computed via the IBM-estimation algorithm.

Work supported in part by a BLN SoLV grant and ASC Honors Undergraduate Research Scholarship.